

⑥ 公開特許公報(A) 平3-135600

⑤ Int. Cl.³

識別記号

庁内整理番号

④ 公開 平成3年(1991)6月10日

G 10 L 9/00
3/00

3 0 1 C
3 0 1 A

8842-5D
8842-5D

審査請求 未請求 請求項の数 1 (全5頁)

⑦ 発明の名称 音声認識装置

② 特 願 平1-273585

② 出 願 平1(1989)10月20日

⑦ 発 明 者 藤 本 潤 一 郎 東京都大田区中馬込1丁目3番6号 株式会社リコー内
⑦ 発 明 者 安 田 晴 剛 東京都大田区中馬込1丁目3番6号 株式会社リコー内
⑦ 出 願 人 株 式 会 社 リ コ ー 東京都大田区中馬込1丁目3番6号

明 示 部 分

1. 発明の名称

音声認識装置

2. 特許請求の範囲

入力する音信号のパワーを検出するパワー検出手段と、検出されたパワーに基づき音声区間を検出する音声区間検出手段と、音声区間検出手段において1つの音声区間を検出後、該音声区間の終端から所定時間以内に次の音声区間が検出される回数を計数し、該計数結果を認識結果として出力する計数手段とを備えていることを特徴とする音声認識装置。

3. 発明の詳細な説明

(産業上の利用分野)

本発明は、特定の複数の単語が音声として入力されたときにこれらを識別しうる音声認識装置に関する。

(従来の技術)

近年、音声認識の分野においては、特に大語彙化、あるいは連続発声された音声を認識させることを指向した研究がなされている。一方、実用上の問題を考えると、現在の音声認識装置では音声を認識させるのに発声に制限をもたせたりしなければならぬので、認識可能な語彙数を多くしても、十分に実用的であるとは言い難い。

実用上の観点からはむしろ、特定の単語だけを認識対象にし、例えば決められた単語を発声しているか否かを判定できるもの、あるいは「ハイ」と「イエエ」を識別できるものが望まれている場合が多い。なお、多くの語彙を認識可能な装置であれば、当然のことながら上記のような2つの単語を識別することもできるが、2つの単語を認識、判定するだけの目的のためには、このような装置は一般に複雑であり、また高コストであった。

ところで、従来、米国特許3688126号には、上記のような2つの単語だけを識別することを意図した実用的な簡易かつ低コストの装置が

開示されている。この装置では、「YES」と「NO」の入力音声とこれらの周波数分布の違いによって識別するようになっている。すなわち「YES」の「S」の音が偏頗音であり高い周波数成分を有していることに着目し、音声区間の終端に高い周波数成分があるか、音声区間全体に高い周波数成分がないかによって、前者を「YES」、後者を「NO」として認識し識別するようになっている。

〔発明が解決しようとする課題〕

しかしながら、「YES」と「NO」とを識別可能な上記音声認識装置を日本語の音声の認識に適用しようとした場合には、日本語の例えば「ハイ」と「イエエ」のいずれも高い周波数成分をもたないで、これらを有効に識別することができない。従って、上記音声認識装置は、英語の音声の識別には適しているが、日本語の音声にはこのままでは利用することができないといった問題があった。

本発明は、特定の日本語の音声を識別するのに

適した実用的な簡易かつ低コストの音声認識装置を提供することを目的としている。

〔課題を解決するための手段〕

上記目的を達成するために本発明は、入力する音声信号のパワーを検出するパワー検出手段と、検出されたパワーに基づき音声区間を検出する音声区間検出手段と、音声区間検出手段において1つの音声区間を検出後、該音声区間の終端から所定時間以内に次の音声区間が検出される回数を計数し、該計数結果を認識結果として出力する計数手段とを備えていることを特徴としている。

〔作用〕

上記のような構成では、パワー検出手段で入力する音声信号のパワーを検出し、検出されたパワーに基づき音声区間検出手段で音声区間を検出するが、識別されるべき特定の複数の単語間でこれらに含まれる破裂音の数が互いに異なっているような場合には、単語全体の発声を終了するまでに検出される音声区間の個数が互いに異なる。すなわち破裂音があると所定時間以内の無音状態が発生

し、単語全体の発声に対して音声区間が区切られる。このことに着目して、計数手段では、1つの音声区間を検出後、該音声区間の終端から所定時間以内に次の音声区間が検出される回数を計数し、この結果を認識結果として出力する。これにより、複数の単語の音声を識別することができる。

〔実施例〕

以下、本発明の一実施例を図面に基づいて説明する。

第1図は本発明に係る音声認識装置の一実施例のブロック図である。

本発明は、日本語では「ハイ」と「イエエ」や「オイ」と「ワルイ」のような対を成す単語として、「マル」(「○」)と「バツ」または「ベケ」(「×」)があることに着目してなされたものであり、第1図を参照すると、本実施例の音声認識装置は、入力音を電気信号に変換するマイクなどの変換部1と、電気信号に変換された音声信号を増幅する増幅部2と、音声信号のパワーすなわちエネルギーを検出するパワー検出部3と、音声

入力される音の周辺雑音のパワーが基準値として記憶される基準値記憶部4と、パワー検出部3で検出されたパワーと基準値記憶部4に記憶されている基準値とを比較し、入力音声の音声区間を検出する音声区間検出部5と、音声区間検出部5で1つの音声区間を検出後、該音声区間の終端から所定の時間以内に次の音声区間が検出される回数を計数し、計数結果に応じた認識結果を出力する計数部6とを備えている。

パワー検出部3は、例えば第2図(a)に示すように、整流回路10と、ローパスフィルタ11とによって実現され、第2図(b)には具体的な回路例が示されている。すなわち増幅部2からの音声信号をダイオードD₁、D₂からなる整流回路10で整流し、抵抗R₁、コンデンサCからなるローパスフィルタ11でフィルタリングすることによって音声信号のパワーを検出するようになっている。また計数部6は、例えばカウンタによって構成されている。

次にこのような構成の音声認識装置の動作例を

説明する。

先づ、対立する単語としての「マル」と「バツ」または「ベケ」の音声を確認し、これらを互いに識別する場合を考える。上記いずれかの単語が発声されると、パワー検出部3で検出された入力音のパワーに基づき、音声区間検出部5では音声区間を検出する。すなわち音声区間検出部5では、パワー検出部3で検出された音信号のパワーが基準記憶部4に記憶されている基準値すなわち周辺雑音のパワーよりも大きくなったときにこれを音声区間の始端として検出する。なお基準記憶部4に記憶される基準値としては、実際の音声が入力するに先立って数ミリ秒間にわたる周辺雑音のパワーデータの平均をとったものが良く、これにより、音声区間の始端を正確に検出することができる。一度音声区間の始端が検出された後は、基準値を更新しないようにし、音声区間の始端を検出後、音声区間検出部5は、音信号のパワーが基準値よりも小さくなったか否かを検知し、基準値よりも小さくなったときにこれを音声区間の終

端として検出する。

ところで、「マル」を発声した場合には、この音声のパワーは第3図(8)に示すようになり、この音声が発声している期間が1つの音声区間 T_0 となる。これに対して「バツ」を発声した場合には、この音声に破裂音「 t 」が含まれ、摩擦音を発するの一度音を閉じるため、後の音「 ts 」の発声前は短かい期間無音状態すなわちパワーが「0」に近い状態となり、これによって第3図(b)に示すように音声区間検出部5では、「バ」を発声している期間と「 ts 」を発声している期間との2つの音声区間 T_1 、 T_2 が検出される。

「ベケ」を発声した場合にも同様に、この音声に摩擦音「 k 」が含まれているので、第3図(c)に示すように「ベ」を発声している期間と「ケ」を発声している期間との2つの音声区間 T_3 、 T_4 が検出される。これら2つの音声区間の間の無音の期間は、後の音が促音である場合や上記のような摩擦音「 t 」、「 k 」、「 p 」であるときに発生するが、促音の場合には無音の期間

は約350ミリ秒程度とされているので、摩擦音の場合にはこれよりもっと短かい時間となる。すなわち上記の例では、最初の音声区間 T_1 、 T_3 の終端が検出された後、約350ミリ秒程度の短かい時間以内に後の音声区間 T_2 、 T_4 の始端が検出されることになる。

計数部6では、1つの音声区間の終端が検出された後、上記時間以内に次の音声区間の始端が検出されたかを監視しており、上記時間以内に次の音声区間の始端が検出されたときには計数値を「1」だけ増進させ、上記時間以内に次の音声区間の始端が検出されなくなった時点で単語全体の発音が終了したと判断し、計数動作を停止してそれまでの計数結果を認識結果として出力する。従って、「マル」が発声されたときには、1つの音声区間 T_0 だけであるので、計数結果は「0」となり、「バツ」や「ベケ」が発声されたときには、後の音声区間 T_2 または T_4 の始端によって計数結果は「1」となる。これにより、極めて簡易な仕方では、「マル」と「バツ」または「ベケ」の対

立する2つの日本語の音声の正確に認識しこれらを識別することが可能となる。

上記説明では、「マル」と「バツ」または「ベケ」を例にとったが、本発明の音声認識装置は、これに限らず、例えば「ゼロ」(「0」)と「イチ」(「1」)のように対立する一方の単語のみ摩擦音を含むものであれば全ての対立する単語の識別に適用することができる。

また2つの日本語の単語の識別のみならず3つあるいはそれ以上の日本語の単語を識別することも可能である。例えば、「マル」(「0」)と、「バツ」または「ベケ」(「 x 」)と、これらの中間を挟む「サンカク」(「 Δ 」)との3つの単語を識別する場合を考える。「サンカク」を発声した場合には、この音声に摩擦音「 k 」が2つ含まれるので第4図に示すように、「サン」を発声している期間と、「カ」を発声している期間と、「ク」を発声している期間との3つの音声区間 T_5 、 T_6 、 T_7 が音声区間検出部5で検出される。

これにより、「サンカク」を発声したときの計数部6の計数結果は“2”となり、計数結果が“0”となる「マル」や計数結果が“1”となる「バツ」または「ベケ」と識別することが可能となる。

(発明の効果)

以上に説明したように本発明によれば、1つの音声区間を検出後、該音声区間の終端から所定時間以内に次の音声区間が検出される回数を計数し、この計数結果を認識結果として出力するようにしているので、簡易かつ低コストの構成で特定の複数の日本語の音声を正確に識別することができる。

4. 図面の簡単な説明

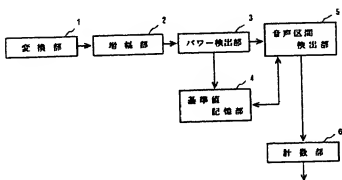
第1図は本発明に係る音声認識装置の一実施例のブロック図、第2図(a)はパワー検出部の構成例を示す図、第2図(b)は第2図(a)に示したパワー検出部の具体例を示す図、第3図(a)、(b)、(c)はそれぞれ単語「マル」、「バツ」、「ベケ」を発声したときの音声のパワーの時間的経過を示す図、第4図は単語「サンカク」を発声したときの音声のパワーの時間的経過を示す図である。

す図、第4図は単語「サンカク」を発声したときの音声のパワーの時間的経過を示す図である。

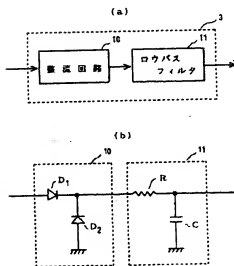
- 1…変換部、2…増幅部、3…パワー検出部、
4…高準位記憶部、5…音声区間検出部、
6…計数部

特許出願人 株式会社 リ コ ー

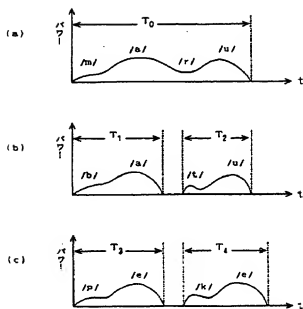
第1図



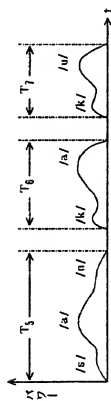
第2図



第 3 図



第 4 図



**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.